

Semantic Permanence in Audiovisual Translation: a FrameNet approach to subtitling

Mairon Morelli Samagaio¹, Tiago Timponi Torrent^{1,2},
Ely Edison da Silva Matos¹, Arthur Lorenzi Almeida¹

¹FrameNet Brasil, Graduate Program in Linguistics, Federal University of Juiz de Fora

²Brazilian National Council for Scientific and Technological Development (CNPq)

mairon.samagaio@letras.ufjf.br; [tiago.torrent|ely.matos]@ufjf.br;

arthur.lorenzi@estudante.ufjf.br

Abstract

This paper presents research on Semantic Permanence of subtitles that translate audiovisual content between two different languages (English and Portuguese). The analysis was made through the semantic annotation of an audio transcription and comparison of the resulting annotation sets with those resulting from the annotation of the subtitles captured from the same video. Our findings indicate that frame semantic cosine similarity between subtitles and audio of the same video can capture the semantic differences between the original spoken sentence and the choices made by the translator to make it possible for the message to fit within the limitations set by the subtitling industry.

1 Introduction

Subtitling is a mode of audiovisual translation affected by a series of restrictions imposed by the industry in which it is inserted. Factors such as spacial/temporal restrictions, and synchrony are expected to create some sort of variation in the semantic pole of the translated sentences that are generated in discourse.

This paper aims to use Frames Semantics (Fillmore, 1982), implemented as an enriched multilingual FrameNet (Torrent et al., 2022), to analyse the semantic permanence of subtitles based on the Primacy of Frame Model (Czulo, 2017). We annotate both the transcriptions of the original audio spoken in English during interview sequences of a TV Travel Series – Pedro Pelo Mundo – aired in Brazil by cable TV channel GNT, and their Brazilian Portuguese subtitles. We then calculate the cosine similarity between the semantic representation which is the result of the annotation tasks for both corpora. We also compare our findings with the ones compiled by Viridiano et al. (2022), which contrasted semantic representations of image captions in English and Brazilian Portuguese (original and the translation from English to Portuguese) be-

tween them, as well as with that of the images they describe in terms of cosine similarity.

Ingestion	[@Action] [@Food] [@Lexical] [#238]
Definition	An Ingestor consumes food or drink (Ingestibles), which entails putting the Ingestibles in the mouth for delivery to the digestive system. This may include the use of an Instrument . Sentences that describe the provision of food to others are NOT included in this
Core Frame Elements	
FE Core:	
Ingestibles	The Ingestibles are the entities that are being consumed by the Ingestor .
Ingestor semantic_type: @sentient	The Ingestor is the person eating or drinking.
Non-Core Frame Elements	
Degree semantic_type: @degree	The extent to which the Ingestibles are consumed by the Ingestor .
Duration	The length of time spent on the ingestion activity.
Instrument semantic_type: @physical_entity	The Instrument with which an intentional act is performed.

Figure 1: The Ingestion frame

This paper contributes to the computational processing of the Portuguese language to the extent that it presents a methodology for calculating the impact of subtitling techniques on the semantics of multimodal data. It also contributes to Frame Semantics by annotating semantic information in multimodal corpora based on two different data sources (audio and subtitles), and analyzing these annotated data in a qualitative way, considering different translation strategies used by professional translators, which also contributes to the field of Translation Studies.

2 Frame Semantics and FrameNet

Frame Semantics and, consequently, its implementation as a FrameNet, is an approach to linguistics studies that, to some extent, emerged in opposition to the then current truth-condition based approaches to meaning (Fillmore, 1985). According to Fillmore (1982), only knowing a word and its definition is similar to having a plethora of utensils at one's disposal, and not knowing what they are used for. As the author goes on about the issue, he affirms that human beings build knowledge from their experiences with the world around them.

Based on this work, Berkeley FrameNet was built as a frame-based lexicon to cover the English language. The database records the following for each frame in it (see Figure 1):

- **Frame Name:** The name that identifies the frame in the database.
- **Frame Definition:** A short definition of the frame, aimed at allowing annotators to identify the main features of the frame in question, as well as the relations between its elements.
- **Frame Elements:** A list of elements that constitute a Frame. Frame elements can be core, or not. For example, in the `Ingestion` frame, the core elements are the `INGESTIBLE`, the substance that is being consumed, and the `INGESTOR`, the entity which is consuming the `INGESTIBLE`. Non-core elements would include, for example, the `INSTRUMENT` used.
- **Lexical Units:** A list of categorized words that evoke the frame in question. For the frame above, two examples would be the verbs `eat` and `devour`.
- **Frame Relations:** Since FrameNet is, in its core, a network of frames, `Ingestion`, is connected to other frames such as `Manipulation` and `Cause_motion` by a series of relations that include inheritance, using, perspective, subframe, among others.

Other FrameNet project have been created for different languages around the world, such as German (Burchardt et al., 2009), Japanese (Ohara et al., 2004), and Brazilian Portuguese (Salomao, 2009). In the last years, research in multimodality has been a topic of growing interest among researchers working with FrameNet (Belcavello et al., 2022;

Viridiano et al., 2022; Torrent et al., 2022; Ciroku et al., 2024). Such is the case of this study, which compares audio and subtitles in two different languages for one audiovisual piece.

2.1 The Primacy of Frame Model

As per Czulo (2013), the Primacy of Frame Model approximates Frame Semantics and Translation. According to the author, the translator's job is to find the maximally comparable frames that must refer to the same scenarios and share core properties in both languages so they can get to the translation of a text.

Czulo (2017) gives an example using the frame for `Marriage` in different cultures: in some, it contemplates a legal stable union between people of different sex. In these cultures, some other frame, like "Partnership", could be applied for a marriage between people of the same sex, since, in the context of that society in particular, the stable legal union between people of the same sex is not legalized or accepted.

2.2 Semantic Permanence

While discussing the Primacy of Frame Model, Czulo (2017) claims that the model is based on the concept that states that when a frame is evoked by a Lexical Unit in a text, all other frames are activated by it, and due to the connection between all frames that form FrameNet, it is possible to see a variation in the frames evoked when a speaker is asked to recreate what they just read or heard.

As in the example given by Petruck (1996), the `Commercial_transaction` frame is associated with different frames, all of them focused on different points of view of it, such as `Commerce_buy`, `Commerce_sell`, or `Cost`. All of these frames are available to the speaker when recreating an experience, and this can lead to variation in the frames evoked by the sentences.

When this variation (or permanence) occurs in different languages, it is called Semantic Permanence. This concept is present in the data analysis of this paper, and it plays a central role in this paper's discussion.

3 Subtitling

Among other reasons, the object of study in this paper - subtitling - was chosen for the limitations imposed by the market to this mode of audiovisual translation.

According to [Cintas and Ramael \(2020\)](#), subtitles are one of the most popular modes of audiovisual translation, and one of its key characteristics is the fact of being added to a final product. This attribute of subtitling creates limitations imposed by the market on the professionals of the area. According to [Cintas and Ramael \(2007, p.145\)](#):

[...]subtitles are limited to two lines, each allowing for a maximum number of characters that cannot be exceeded, depending on the time the subtitle remains on screen[...]. This is why traditional commercial subtitling has developed a style of its own that has an impact on grammar and register, as well as on the interactional and other oral features of dialogue.

The limitations of subtitling are threefold, and can be categorized as follows:

- **Spatial limitations:** A subtitle must not be longer than two lines, and cannot be bigger than 1/12 of the screen, this measure is estimated to accommodate approximately 42 characters.
- **Temporal limitations:** subtitles must not stay on display for longer than six seconds. Following the information above, a subtitle with two lines of 42 characters, that must stay in display for no longer than six seconds has a limitation of about 14 characters per second. However, evolving conventions on the industry have turn this number to 17 characters per second, calculating that this makes an average of 200 words per minute.
- **Synchrony:** Synchrony is one of the most important characteristics of subtitling. It dictates that the subtitles must not speed up or down any information of the audiovisual piece so as not to impact the viewer’s perception of the video.

As a result of the limitations above, it is safe to affirm that the product of subtitling tends to be a reduction of the original text present in the audio of the translated information.

The question of what information must be reduced or maintained in a text is determined by the situation. Depending on the context, translators may decide to keep some information or reduce it in some way, and the strategies used to reduce

information may vary from a simplification of the discourse to complete omission of the information in question (see also [Cintas and Ramael \(2020\)](#) for a more detailed discussion on the subject).

4 Materials and methods

As we aim to capture the differences between audio and subtitles in a multimodal dataset, we worked with a dataset of 951 comparison sets¹ of original English audio transcription and Brazilian Portuguese subtitles for the experiment reported on in this paper.

The comparison sets were extracted from the Pedro Pelo Mundo corpus². This corpus comprises ten episodes of a TV travel series in which the host travels to different countries and interviews locals. When the person being interviewed does not speak Portuguese, English is used, and there are subtitles to translate the interview into Brazilian Portuguese.

While creating the Pedro Pelo Mundo corpus, all ten episodes of the documentary series were treated by the Charon pipeline ([Belcavello et al., 2022](#)), in which the video files pass through a speech-to-text algorithm in which all the spoken data present in the video is transcribed, and further revised by human annotators. Also, a text recognition program captures the subtitles present on screen. Subtitles are also revised by human annotators.

The product of this process was two subcorpora – one for original English audio and one for Brazilian Portuguese subtitles – whose statistics are presented in Table 1:

	En audio	Br-Pt subs
Tokens	13,052	9,366
Words	10,916	7,907
Sentences	1,717	1,743
Documents	1	1

Table 1: Subcorpora statistics

Once all text data is captured by the pipeline, all the data was proofread by native and fluent speakers of the languages present in the subcorpora. Subsequently to these processes, the sentences are annotated following the full-text annotation method-

¹The term comparison set was chosen because there are cases where one sentence from the audio is broken into two or more different subtitles after the corpora were aligned.

²The Pedro Pelo Mundo corpus is also being used in other researches carried out in FrameNet Brasil at this moment, so the experiments done here can also cooperate with further findings on the works previously quoted.

ology devised by FrameNet (Ruppenhofer et al., 2016).

Figure 2: Annotated sentence in Brazilian Portuguese

Figure 2 presents an example of a fully-annotated sentence in Brazilian Portuguese. The annotation process tags semantic and syntactic properties of the sentences, which is a fundamental part of our analysis, as the annotation sets are the main information that are compared between languages.

Once fully annotated, the subcorpora are ready to go through an application of a spread activation technique that calculates similarities based on *soft* cosine similarity. The spread activation algorithm is a program that allows for measuring the differences or similarities in frames evoked by two or more different sentences. Viridiano et al. (2022) use the same method to make comparisons between sentences, and elucidate how the process works:

The SA algorithm models an iterative energy propagation process from one or more nodes to other nodes in a graph in three stages: (i) pre-adjustment, (ii) spreading, and (iii) post-adjustment[...]. Before the spreading stage, the energy value for each node was calculated during the pre-adjustment stage. Energy decay was calculated for the value of the node so that this value is within the [0,1] interval. The calculated value was then output to the neighboring nodes. Post-adjustment was not used, since the FN graph is acyclic and the FN hierarchies do not comprise many levels. (Viridiano et al., 2022, p.111)

This method assigns higher scores of cosine similarity to more similar annotation sets, and lower

scores to less similar annotation sets, based on the distance which the algorithm had to go through to get from a frame (in the original text) to another frame (in the translated text).

As an example, consider the sentence pair in (1a) and (1b). Although (1b) is a translation of (1a), the frames evoked in the two sentences are not exactly the same. The words *crisis.n* and *crise.n*, in English and Portuguese, respectively, evoke the Catastrophe frame, which is the main one for both sentences. However, the fact that *big.a* evokes the Size frame in the English original does not find a correspondence in the translation.

- (1) a. There was a big crisis on Denmark in the time.
- b. A Dinamarca estava em crise naquela época.

The variation between the sentences can be depicted by means of a graph showing the relations between the frames evoked and as well as other frames in the network (Figure 3).

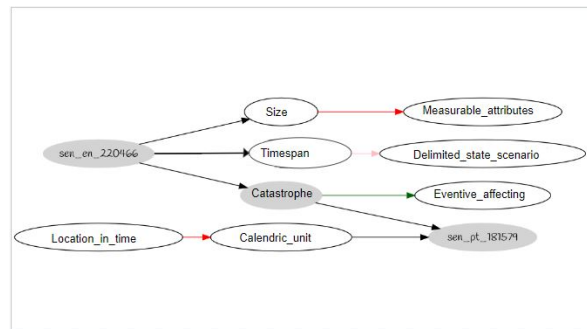


Figure 3: Graph representing the frames evoked by the sentences in (1a-1b)

5 Quantitative Data Analysis

As previously mentioned, the experiment reported here gathered 951 comparison sets that are a result of the alignment of the corpora. We separated the data collected into two groups: original, and translation. The first method of analysis of corpora chosen for this paper was the Student's t-test, since it is able to compare datasets with two distinct origins (Lopes et al., 2015).

5.1 Student's t-test

We conducted two separated tests: the first one took into consideration all the data, including sentences from the audio which were erased during the

translation, while the second one only took into account the sentences that had received a translation in the subtitles. This separation in two tests aims to analyze if there was semantic information in the sentences erased in the translation. The results are presented on Figures 4 and 5.

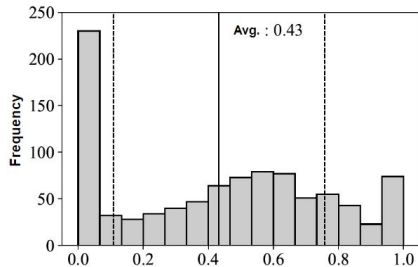


Figure 4: Student's t-test result for all sentences

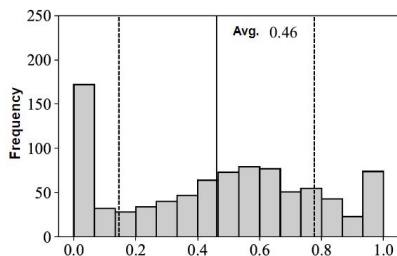


Figure 5: Student's t-test result comprising only the sentences that have subtitles

For the first comparison set (Figure 4), the cosine similarity score is 0.43, and the standard deviation rate is 0.32. For the second comparison set (Figure 5), the cosine similarity rose to 0.46 and the standard deviation rate fell to 0.31. Beyond that, other information must be taken into account here. The t-test statistics was $(892) = 2665$ and $p\text{-value} = 0.007$. It is possible to affirm, by analyzing the data obtained by the Student's t-test, that the sentences that were not subtitled had semantic information, and thus, the absence of this information in the comparison set featuring all of the sentences resulted in lower semantic similarity, indicating less semantic permanence between original and subtitle.

5.2 Comparison with Previous Research

Viridiano et al. (2022) conducted a study comparing the frame-based annotation of images and descriptions in the Flickr30k (Young et al., 2014) dataset and its extensions: Multi30k (Elliot et al., 2016) and Flickr30k Entities (Plummer et al., 2015).

The authors analyze the difference between (i)

frames evoked by static images and descriptions of those images in English, (ii) the original English descriptions and their translations to Brazilian Portuguese, and (iii) the original English descriptions for the images and original Brazilian Portuguese descriptions produced for the same images Viridiano et al. (2022).

The results found by Viridiano et al. (2022) show a cosine similarity of 0.51 by comparing the original English descriptions to their respective translations in Brazilian Portuguese. When the comparison took into account the original descriptions produced for a given image in both languages, the cosine similarity between semantic annotations was 0.33. Finally, while comparing the original descriptions in English with the frames evoked by the image annotation, the cosine similarity received a score of 0.43. This result is similar with the score found in this research for the comparison between the annotated corpus based on all the sentences (translated, or not) and the subtitle corpus.

A possible framing of these results points to the conclusion that the frame permanence of subtitles is similar to that of intermodal translation. In other words, the cosine similarity found taking into consideration the whole set of audio sentences and the subtitles is similar to the one found between the frame annotation of images (visual mode) and the descriptions (verbal language mode) accompanying them in contrast to the translation of the descriptions and the frames evoked by the images themselves.

6 Qualitative Data Analysis

After the corpora were created, annotated, analyzed, and compared for cosine similarity, it was possible to further improve our analysis in a qualitative approach. For this section of the paper, we focus on emblematic cases which help translate cosine similarity values into real data examples.

6.1 Full Semantic Similarity

The first case is that of complete semantic similarity between original and translation. In the comparison dataset, there were a total of 60 cases of full semantic similarity. These cases are representative of translations in which there is a convergence of the frames evoked by both original – e.g. (2a) – and translation – e.g. (2b).

- (2) a. How many records do you have here?
 b. Quantos discos você tem aqui?

In both sentences, the frames evoked were `Records`, by the Lexical Units (LUs) *records.n* and *discos.n*, `Possession` by the LUs *ter.v* and *have.v*, and `Relative_relation` by the LUs *here.adv* and *aqui.adv*.

As it is possible to observe here, any divergences or alterations that may be present as an outcome of the translation between languages, or from the audio to the subtitle, do not affect the frames evoked by both sentences. The translation respects the Primacy of Frame Model, by maintaining the same frames evoked in both sentences.

6.2 Null Semantic Similarity

At the opposite end of our spectrum, we have cases where the cosine similarity between the semantic annotations of sentences is 0,00. These cases represent 217 cases of our total of 951 comparison sets. It was possible to further divide those cases into three different subcases: (i) total divergence of evoked frames (38 cases); (ii) sentences erased by the translator (54 cases); and (iii) sentences that do not evoke frames (125 cases).

6.2.1 Lack of Shared or Related Frames

This is the case in which one of the sentences spoken by a participant of the show has been translated as two different lines of subtitles that had no shared or related frames, as seen in (3a-3b).

- (3) a. That's a good point, so I will have the whale.
 b. Tem razão. Vou querer uma então.³

The frame-evoking LUs in the English sentence are: *good.a*, evoking the `Desirability` frame; *point.n*, evoking the `Topic` frame; *have.v*, evoking the `Ingestion` frame, and *whale.n*, evoking the `Ingredients` frame.

On the other hand, the sentences in Brazilian Portuguese comprise fewer LUs and, consequently, fewer frames, due to the subtitling strategy used in this case. The LUs that evoke frames are *razão.n*, evoking the `Reason` frame, and *querer.v*, evoking the `Desiring` frame.

The decisions made by the translator have altered the sentences in a way that no frame convergence

³You have reason So will have one (literal translation of the sentence in Brazilian Portuguese)

can be found between them, therefore leading to a score of 0 for this example.

6.2.2 Sentences that do not Evoke Frames

This is the case where a sentence does not evoke a frame in one or both languages. For this case, we chose an example of sentence that does not evoke frames for any of the languages.

- (4) a. You have to taste first, and then smell it.
Okay.
Allright.
 b. Coma antes de cheirar⁴.
Está bem.
Vamos lá.

For the example (4a), the second and third sentences do not evoke any frames, since they have a more pragmatic function in the fragment.⁵ Therefore, it is impossible to have a convergence of frames for the comparison set.

6.2.3 Sentences Erased by the Translator

In this case, the translator chose not to include some sentence in the subtitles as a strategy to cope with time and space limitations.

- (5) a. That's bad?
 No, it's delicious.
 Delicious? Okay.
You're pretty sure that
 It's kind of like how they named Iceland.
 b. É ruim?
 Não, é delicioso.
 Delicioso? Está bem.

É parecido com o jeito que nomearam a Islândia⁶.

In this fragment - (5a) -, comprised by two questions made by the interviewer and the respective answers given by the interviewee, the fourth sentence does not have a correspondent translation in (5b). Therefore, the similarity found in this case is 0.

⁴Eat before you smell it (literal translation of the sentence in Brazilian Portuguese)

⁵The current implementations of FrameNet do not cover pragmatic phenomena extensively. For a discussion see Czulo et al. (2020).

⁶It is kind of like how they named Iceland (literal translation of the sentence in Brazilian Portuguese)

6.3 Average Semantic Similarity

The comparison sets showing average semantic similarity correspond to a total of 324 pairs, whose cosine similarity scores ranged from 0.41 to 0.69.

A qualitative analysis of those sets allows us to find all the strategies compiled by Cintas and Ramael (2020) which indicates choices translators have to make to overcome time and space limitations imposed to subtitling. In some of the cases, depending on the context, the translator used up to two strategies at the same time. For this paper, we chose three different cases of average cosine semantic similarity as examples.

6.3.1 Omission and Generalization of Enumerations

In (6a-6b), it is possible to see the use of two strategies by the translator when a sentence has too much information to be accommodated in the subtitles, given the limitations imposed by the industry.

- (6) a. They are really peaceful until you try to make a road through their elf city, or build a house.
- b. Eles são pacíficos até você tentar construir algo em suas terras ⁷.

The linguistic choices made by the translator in their strategy were the omission of the intensifier *really.adv* and the generalization of enumerations, by replacing “make a road” and “build a house” with “construir algo” (*to build something*). These choices made a substantial difference in the frames evoked in both sentences.

The English sentence evokes the following frames: Degree, Personality_traits, Time_vector, Attempt, Roadways, Traversing, Locale_by_use, Building, Buildings. In Brazilian Portuguese, the frames evoked by the sentence are: Personality_traits, Attempt, Building, Entity and Political_locales.

Although the choices made by the translator represent a substantial difference between the sentences, it is possible to see a concern in using frames that are close to each other. Since FrameNet frames are interrelated (Figure 6), it is possible to use the network structure to compare the two

⁷They are peaceful until you try to build something in their land (literal translation of the sentence in Brazilian Portuguese)

sentences, showing that their differences are not enough to give a score of 0 to the comparison set.



Figure 6: Graph representing the frames evoked by the sentences in (6a-6b)

6.3.2 Simplification and Alteration at the Sentence Level

In (7a-7b) we have an example of a sentence pair in which the translator used two of the strategies compiled by Cintas and Ramael (2020): the simplification of the sentence and the alteration at the sentence level.

- (7) a. We have, in our mind, always to help the people.
- b. Nossa cultura sempre foi muito solidária⁸.

The first aspect of this translation that can be highlighted is the alteration of “We have in our mind” to “Nossa cultura (Our culture)”. Also, the verb phrase “to help the people” was substituted by the simpler noun phrase structure “muito solidária (very solidary)”. In this context, the interviewer asks the interviewee a series of questions about Myanmar’s opening to the world. The main information is on how the people reacts to tourists coming to the country.

In the Frame Semantics pole of our analysis, we can see that the frames in English are Possession, Body_parts, Frequency, Assistance and People. In the subtitle, the frames are Fields, Frequency and

⁸Our culture was always very solidary (literal translation of the sentence in Brazilian Portuguese)

Attributes. These changes in the frames evoked by the translated sentence were enough to give the comparison set a total score of 0.5.

6.3.3 Simplification of Verb Tenses

In the excerpt in (8a–8b), it is possible to see yet another strategy a translator can use while creating subtitles for audiovisual translation (Cintas and Ramael, 2020): the use of simpler verb tenses.

- (8) a. What would you say is the best thing about Singapore?
b. O que há de melhor em Singapura?⁹

The original sentence evokes the frames *Statement*, *Desirability*, and *Entity*. However, the translated text evokes another set of frames: *Existence* and *Desirability*. This is a consequence of the simplification of the conditional structure to an existential one in the present tense.

Even though the difference in semantics and syntax is considerable, it is not able to give a score of 0 to the comparison sets, assigning a score of 0.5 to it, given the existence of relations between frames.

7 Discussion

As per the examples analyzed in the previous section, it is possible to find the semantic differences in the translation caused by the strategies listed by Cintas and Ramael (2020), which were used by translators in response to the limitations imposed to subtitling by the industry. Such limitations impact the frame semantic cosine similarity between the original sentence and the translation. The use of the frame-based cosine similarity score allows for keeping track of and classifying the different impacts of translators' choices mathematically.

The proposed metric allows for the comparison between original audio and subtitle translation even when the frames evoked by each sentence are completely different. This is so because the implementation of the metric relies on a spreading activation technique on the FrameNet network of frames. Therefore, we believe that this research contributes to the Primacy of Frame Model as postulated by Czulo (2017), by providing a means to measure the notion of Maximal Comparability.

It was also possible to show, based on the data analyzed in this paper, that the differences between

⁹What is best in Singapore?(literal translation of the sentence in Brazilian Portuguese)

the original and the translation are not caused only by the systemic differences between languages, but also by the strategies used by the translator during the adaptations needed to respect the spatial and temporal limitations imposed by the industry onto subtitling.

8 Conclusion

This paper presents a metric for analysing the maximal comparability between source and translated texts (Czulo, 2017) based on Frame Semantics (Fillmore, 1982).

The application of this metric to a corpus featuring original English audio and Brazilian Portuguese subtitles showed that subtitles are closer to an intermodal translation (Viridiano et al., 2022) than to a translation of a written text to a different written text in another language, not just by the difference caused due to the adaptation from spoken speech to written text, but also because of the adaptations necessary to reach the standards created by the industry (Cintas and Ramael, 2007) not to undermine the understanding of the original.

The technique chosen for the metric, namely spread activation, was able to capture the differences in the semantic pole of the data, leading to relevant conclusions on subtitling as a modality of audiovisual translation, allowing for an analysis on the comparison level of both, *corpora* and sentence with results comparable to previous research on the area (Viridiano et al., 2022).

9 Acknowledgments

Authors acknowledge the support of the Graduate Program in Linguistics at the Federal University of Juiz de Fora. The FrameNet Brasil Lab is funded by FAPEMIG grant n° RED 00106-21, and by CNPq grants n° 408269/2021-9 and 420945/2022-9. Samagaio's research presented in this paper was funded by CAPES PROEX grant n° 88887.816242/2023-00. Torrent is an awardee of CNPq's Research Productivity grant 315749/2021-0.

References

- Frederico Belcavello, Marcelo Viridiano, Ely Matos, and Tiago Timponi Torrent. 2022. *Charon: A FrameNet annotation tool for multimodal corpora*. In *Proceedings of the 16th Linguistic Annotation Workshop (LAW-XVI) within LREC2022*, pages 91–96, Marseille, France. European Language Resources Association.

- Aljoscha Burchardt, Katrin Erk, Anette Frank, Andrea Kowalski, Sebastian Padó, and Manfred Pinkal. 2009. Framenet for the semantic analysis of german: Annotation, representation and automation. *Multilingual FrameNets in Computational Lexicography: methods and applications*, 200:209–244.
- Jorge Diaz Cintas and Aline Ramael. 2007. *Translation Practices Explained*. Routledge, Oxfordshire, England, UK.
- Jorge Díaz Cintas and Aline Ramael. 2020. *Subtitling: Concepts and Practices*, 1 edition. Translation Practices Explained. Routledge, Oxfordshire, England, UK.
- Fiorela Ciroku, Stefano De Giorgis, Aldo Gangemi, Delfina S. Martinez-Pandiani, and Valentina Pre-sutti. 2024. [Automated multimodal sensemaking: Ontology-based integration of linguistic frames and visual data](#). *Computers in Human Behavior*, 150:107997.
- Oliver Czulo. 2013. Constructions-and-frames analysis of translations: The interplay of syntax and semantics in translations between english and german. *Constructions and Frames*, 5(2):143–167.
- Oliver Czulo. 2017. Aspects of a primacy of frame model of translation. *Empirical modelling of translation and interpreting*, 7:465.
- Oliver Czulo, Alexander Ziem, and Tiago Timponi Torrent. 2020. [Beyond lexical semantics: notes on pragmatic frames](#). In *Proceedings of the International FrameNet Workshop 2020: Towards a Global, Multilingual FrameNet*, pages 1–7, Marseille, France. European Language Resources Association.
- Desmond Elliot, Stella Frank, Khalil Siman, and Lucia Specia. 2016. [Multi30K: Multilingual English-German image descriptions](#). In *Proceedings of the 5th Workshop on Vision and Language*, pages 70–74, Berlin, Germany. Association for Computational Linguistics.
- Charles J. Fillmore. 1982. Frame semantics. In The Linguistics Society of Korea, editor, *Linguistics in the Morning Calm*, pages 111–137. Hanshin Publishing Co., Seoul, South Korea.
- Charles J. Fillmore. 1985. [Frames and the semantics of understanding](#). *Quaderni di Semantica*, 6(2):222–254.
- Aline Cristina Berbet Lopes, Amanda da Cruz Leinioski, and Larissa Ceccon. 2015. Testes t para comparação de médias de dois grupos independentes. *Universidade Federal do Paraná–UFPR–Departamento de Zootecnia*.
- Kyoko Hirose Ohara, Seiko Fujii, Toshio Otori, Ryoko Suzuki, Hiroaki Saito, and Shun Ishizaki. 2004. The japanese framenet project: An introduction. In *Proceedings of LREC-04 Satellite Workshop “Building Lexical Resources from Semantically Annotated Corpora” (LREC 2004)*, pages 9–11.
- Miriam R.L. Petruck. 1996. Frame semantics. In Jef Verschueren, Jan-Ola Östman, Jan Blommaert, and Chris Bulcaen, editors, *Handbook of Pragmatics*. John Benjamins, Amsterdam, NE.
- Bryan A. Plummer, Liwei Wang, Chris M. Cervantes, Juan C. Caicedo, Julia Hockenmaier, and Svetlana Lazbenik. 2015. [Flickr30k entities: Collecting region-to-phrase correspondences for richer image-to-sentence models](#). In *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 2641–2649.
- Josef Ruppenhofer, Michael Ellsworth, Myriam Schwarzer-Petruck, Christopher R Johnson, Collin F. Baker, and Jan Scheffczyk. 2016. *FrameNet II: Extended Theory and Practice*. International Computer Science Institute, Berkeley, CA.
- Maria Margarida Martins Salomao. 2009. Framenet brasil: um trabalho em progresso. *Calidoscópio*, 7(3):171–182.
- Tiago Timponi Torrent, Ely Edison da Silva Matos, Frederico Belcavello, Marcelo Viridiano, Maucha Andrade Gamonal, Alexandre Diniz da Costa, and Mateus Coutinho Marim. 2022. [Representing context in framenet: A multidimensional, multimodal approach](#). *Frontiers in Psychology*, 13.
- Marcelo Viridiano, Tiago Timponi Torrent, Oliver Czulo, Arthur Lorenzi Almeida, Ely Edison da Silva Matos, and Frederico Belcavello. 2022. [The case for perspective in multimodal datasets](#).
- Peter Young, Alice Lai, Micah Hodosh, and Julia Hockenmaier. 2014. From image descriptions to visual denotations: New similarity metrics for semantic inference over event descriptions. *Transactions of the Association for Computational Linguistics*, 2:67–78.